Metadata Recommendations Supporting Data Discovery and Use in Data.gov and the Geospatial Platform

June 14, 2017 update

Contents

Introduction

Metadata creation and management is a geospatial data management best practice. When geospatial data resources are documented robustly and in accord with national and, preferably, international, metadata standards, the resulting metadata record enhances the ability of the resource to be:

- discovered
- assessed
- accessed
- applied
- archived

When well-documented metadata is published to Data.gov it become discoverable using keywords, time period, geographic extent, organization name, and other search criteria. The Geospatial Platform extends the use of the Data.gov metadata by providing applications for download, visualization, data compilation, and custom map production.

This document identifies best practices for creating metadata records that support the full range of Data.gov and Geospatial Platform operational capabilities and improve the discovery and use of geospatial data resources. The document also serves as a foundation for companion

metadata guidance developed for National Geospatial Data Assets (NGDA) and Project Open Data.

- National Geospatial Data Assets Metadata Guide
- Project Open Data implementation Guide

FGDC-Endorsed Metadata Standards

The Federal Geographic Data Committee authored and endorsed the federal Content Standard for Digital Geospatial Metadata (CSDGM) and many agencies and non-federal organizations have utilized the CSDGM since its publication in 1994 and revision in 1998. Since that time, the US standards community has participated in International Standards Organization (ISO) efforts to develop geospatial metadata standards that address newer data formats and content and apply to a broader range of geospatial data resources (services, data models, etc.). Given that federal policy OMB A119 Revised calls for agencies to use voluntary consensus standards, such as those developed through ISO, in lieu of government-unique standards, the FGDC now endorses several of the ISO 191** series of geospatial standards. Agencies publishing to Data.gov and the Geospatial Platform are therefore encouraged to create metadata using:

- *ISO 19115: Geographic Information – Metadata* content standard and formatted using the *ISO/TS 19139:2007: Geographic information -- Metadata -- XML schema*

Data.gov efforts are underway to also support the ISO 19115-1 publication of metadata created using the recent update to ISO 19115:

- *ISO 19115-1:2014: Geographic information -- Metadata -- Part 1: Fundamentals* and formatted using *ISO/TS 19115-3:2016: Geographic information -- Metadata -- Part 3: XML schema*

Metadata Content Best Practices

Write Informative Titles and Abstracts

A good title is descriptive and distinctive. It provides consumers a good sense of the resource content and context and enables them to distinguish among similar resources. *Titles* should not try to replace an *Abstract* or *Purpose* statement, but they should strive to relay the *what, when, where* and, if relevant, *who, why,* and *how* of the resource.

- What is the resource topic?
- When did the content occur or when was it captured?
- Where is the content located on the earth?
- Who is the authority and source for the resource?
- How is the resource formatted?
- Why was the resource created?

Title example:

Aquifer Systems and Recharge Potential in Louisiana from Louisiana DEQ source data, Geographic NAD83, Louisiana Oil Spill Coordinator's Office (LOSCO), 1999, [aqrgeog3dpdeq]

National Geospatial Data Asset (NGDA) data managers should reference the NGDA Metadata Guidelines for additional information about best practices for titles that distinguish NGDA resources.

A good abstract expands upon the title and enables users to better assess the data resource's fitness for use. The abstract should include:

- a general description of the data resource content and features
- form of the data resource, e.g. GIS, imagery, database, service, application, etc.
- relevant place names and references
- the time period of the data resource content
- information about special data characteristics or limitations, e.g. data access limitations, excluded geographies or content, etc.

Specify Dates Using the Correct Format

Both CSDGM and ISO metadata provide multiple date elements for both the metadata and the data resource. These include metadata creation, publication, and update dates; data resource creation, publication, update, content, and processing dates; and source publication and content dates.

If using ISO 19115/-1:

The *ISO 8601* date format of YYYY-MM-DD is recommended. If the exact day or month is not known, the convention allows for the use of YYYY-MM and YYYY.

If using CSDGM:

The format YYYYMMDD, is required. If the exact day or month is not known, the convention allows for the use of just the year, YYYY. The 6-letter representation, YYYYMM, e.g. 201112 (December of 2011), is not recommended as it is easily misinterpreted as the incorrect, but still used, YYMMDD (November 12, 2020).

Each of the standards allows the use of text to represent indeterminate dates. In the CSDGM, "unknown' is an acceptable value and "present" is an acceptable end-date for a continuously collected data resources. In ISO, an indeterminate value of "unknown", "after", "before", and "now" can be used.

Use Unique Identifiers

Unique identifiers are character strings associated with a single entity. Identifiers increase the efficiency and accuracy of metadata. ISO metadata provides unique identifiers as options for many ISO metadata elements, especially in citations where they can be used to uniquely identify both data and metadata resources. Identifiers are also used to reduce the unnecessary restatement of standard parameters for commonly used:

- Instruments and Platforms (e.g. NOAA)
- Spatial Reference Systems (e.g. EPSG)
- Geographic Locations (e.g. GNIS)

Identifiers should be created for both the dataset resources (data or service) being documented and, where possible, the resource (dataset) metadata record.

If using ISO 19115:

The resource identifier should be included as the resource Citation Identifier:
MD_Metadata.identificationInformation>MD_Identification.citation>CI_Citation.identifier

The metadata record identifier should be included as the Metadata Identifier:
*MD_Metadata.*fileIdentifier

If using ISO 19115-1:

The resource identifier should be included as the resource Citation Identifier:
MD_Metadata.identificationInformation>MD_Identification.citation>CI_Citation.identifier

The metadata record identifier should be included as the Metadata Identifier:
*MD_Metadata.*metadataIdentifier

If using CSDGM:

The resource identifier should be included as the Other Citation Details
metadata>idinfo>citeinfo>othercit
and formatted within the element as: resourceIdentifier="(*identifier value*)"
e.g, **Other Citation Details:** resourceIdentifier= "doi:10.7358/V5057CV7"

The metadata record identifier should be included as the Other Citation Details:
metadata>idinfo>citeinfo>othercit
and formatted within the element as: metadataIdentifier="(*identifier value*)"
e.g, **Other Citation Details:** metadataIdentifier= "gov.noaa.ngdc.stp.terrestrial:G37468"

Identifiers are best represented using a Unique Resources Identifier (URI) - an IETF/W3C standard scheme for treating resource identifiers as globally unique and persistent, even when the resource ceases to exist or becomes unavailable. URIs are used to unambiguously identify an abstract or physical resource (e.g., a place, standard, role, namespace, dataset, or service) and are especially useful when searching for resources or things related to them. Part of a URI defines the namespace of the identifier. For example, the URI "EPSG:4326" is an abbreviated identifier of a Coordinate Reference System specification defined by the European Petroleum Survey Group, a well-known naming authority for geodesy parameters. Note: Uniform Resource Locators (URLs) are a common form of URI but are notoriously unreliable resource identifiers as the network location of resources frequently change. [see URI reference[1]]

When generating URIs for a data resource or metadata record, it is highly recommended to use globally unique URIs that are either:

- issued by a well-established organization that maintains a registry of identifiers such as the DOI Foundation (http://doi.org/)

or

- developed using a hashing algorithm with properties that ensures that it is globally unique.

Additional information about unique identifiers is available from:
https://en.wikipedia.org/wiki/Universally_unique_identifier.

[1] [URI reference]: IETF RFC 2396 (August 1998), Uniform Resource Identifiers (URI): Generic Syntax, Berners-Lee, T., Fielding, N., and Masinter, L., eds., <http://www.ietf.org/rfc/rfc2396.txt>

Provide Rich Keywords and Use Controlled Vocabularies

Data.gov and the Geospatial Platform rely heavily on keywords to identify resources of highest interest to users. Keywords of all types should be provided, as relevant, including:

- Theme
- Place
- Stratum
- Temporal
- Discipline, Instrument, process, project, etc. (ISO options)

Keywords should be listed individually and not grouped into delimited lists, e.g. 'wetlands, marsh, salt-marsh, bogs'.

ISO metadata requires the selection of one or more *Topic Categories* that best characterize your data or application domain. These are very high-level thematic bins used by Data.gov and the Geospatial Platform to quickly sort available resources. For those still using the CSDGM standard, include one or more *Topic Categories* as a *Theme Keyword* and specify "ISO 19115 Topic Category" as the *Theme Keyword Thesaurus*.

ISO 19115-1 Topic Categories

biota	*intelligenceMilitary*
boundaries	*location*
climatologyMeteorologyAtmosphere	*oceans*
economy	*planningCadastre*
elevation	*society*
environment	*structure*
farming	*transportation*
geoscientificInformation	*utilitiesCommunication*
health	*extraTerrestrial*
imageryBaseMapsEarthCover	*disaster*
inlandWaters	

Keywords are most effective when drawn from theme and discipline-specific controlled vocabularies such as the *USFWS Cowardin Wetland Classification System* taxonomy, the *Global Change Master Directory Science Keywords*, and the *Global Names Index Service (GNIS)* gazetteer. The use of controlled vocabularies helps to standardize both the spelling of the term and the definition and therefore improves the relevancy of the search results. When documenting keywords from controlled vocabularies, provide:

- the name of the vocabulary as the *Thesaurus Name*
- a direct URL to the controlled vocabulary as an *Online Linkage* within the thesaurus *Citation,* and
- if using ISO and if available provide a URI for the controlled vocabulary as the thesaurus *Citation.identifier*.

If the vocabulary is not published, provide a contact for the vocabulary and encourage the contact to make the vocabulary available online.

National Geospatial Data Asset (NGDA) data managers should reference NGDA Metadata Guidelines for additional information about "Theme Keywords" that distinguish NGDA resources.

Include Links to Browse Graphics

Both CSDGM and ISO provide for the reference of a browse, or "thumbnail" graphic. Browse graphics are especially valuable for the identification of relevant search results. A picture tells a thousand words and a quick view of the data can often provide sufficient context for a user to discern its appropriateness. If a URL for the graphic is placed in the "Browse Graphic" metadata element, Data.gov and the Geospatial Platform can display the graphic with the search results

Provide Reliable Contact Information

All organizations experience staff changes. Contact information should therefore focus on the organization and include the *Agency Name, Position Name, Agency Email* and *Phone*. Individual contact information should be provided as relevant/able. The record should include a *Point of Contact* for the resource being documented, a *Metadata Contact*, and contacts for resource *Distribution* and *Processing* as relevant. The use of ISO *Responsible Party Role Codes* is highly recommended. The role codes make clear the relationship of the contact to the resource, e.g. owner, distributor, processor, principal investigator, etc.

Provide Direct URLs to Online Resources

Resources (data, services, applications) that are described in either CSDGM or ISO metadata must include URLs that take a user to the online resource. This is a baseline expectation of Data.gov. **URL links must provide direct access to the resource wherever possible**. Metadata records that only include references to websites or HTML pages where one must re-initiate search are not user- friendly and preclude automated clients such as the Geospatial Platform from accessing the data. As both metadata standards allow for multiple resource links to be placed in metadata, the following guidance is provided to facilitate users to directly access multiple facets of the data: data download, web services, and data documentation.

Data Download URLS

If using the CSDGM, the data download URL should be documented in the "Distribution" section of the metadata record and provided within the "Digital Transfer Online Network Resource Name" element. In addition the value 'download' should be provided within the "Distribution Digital Transfer Format Specification" element. The "Citation – Online Linkage" should not be used to designate the download URL as it does not provide supporting elements to make clear the download function of the URL.

If using ISO, the data download URL should be documented as a "Citation – Online Linkage". The "Citation – Online Linkage" can originate from the "Identification" or "Distribution" section. The format of the download should be documented in the "Citation – Online Resource Application Profile' and the value "download" should be provided within the "Citation – Online Resource Function" element.

Web Service URLS

Metadata for a web service should include:

1. An *actionable* (i.e., online and consumable) service endpoint URL that provides direct access to the geospatial web service of the specified resource type. The initial emphasis is to enable uniform and reliable access to a dataset as maps and layers via online services that are compliant with the OGC WMS and/or Esri REST API specifications. If there are multiple services for an individual dataset, all the endpoint URLs for map services that host the dataset should be documented.

2. The URI that uniquely identifies the appropriate service application profile specification[2] associated with the geospatial web service:
   - **OGC Web Map Service (WMS)**
     - Specification Identifier: **http://opengis.net/spec/wms** – a service compliant with an approved OGC Web Map Service implementation specification, or specific version, e.g., OGC Web Map Service version 1.1: http://opengis.net/spec/wms/1.1
   - **OGC Web Feature Service (WFS)**
     - Specification Identifier: **http://opengis.net/spec/wfs** – a service compliant with an approved OGC Web Feature Service implementation specification, or specific version, e.g., OGC Web Feature Service version 1.0: http://opengis.net/spec/wfs/1.0
   - **OGC Web Coverage Service (WCS)**
     - Specification Identifier: **http://opengis.net/spec/wcs** – a service compliant with an approved OGC Web Coverage Service implementation specification, or specific version, e.g., OGC Web Coverage Service version 1.0: http://opengis.net/spec/wcs/1.0
   - **OGC Web Map Tile Service (WMTS)**
     - Specification Identifier: **http://opengis.net/spec/wmts** – a service compliant with an OGC Web Map Tile Service implementation specification, or specific version, e.g., OGC Web Map Tile Service 1.0.0: http://opengis.net/spec/wmts/1.0.0
   - **OGC Catalog Service (CSW)**
     - Specification Identifier: **http://opengis.net/spec/csw** – a service compliant with an OGC Catalog Service for the Web implementation specification, or specific version, e.g., OGC Catalogue Service Specification 2.0.2: http://opengis.net/spec/csw/2.0.2
   - **OGC Keyhole Markup Language (KML)**
     - Specification Identifier: **http://opengis.net/spec/kml** – a service that produces a document that is compliant with the OGC Keyhole Markup Language specification, or specific version, e.g., OGC Keyhole Markup Language version 2.2: http://opengis.net/spec/kml/2.2

[2] The Open Geospatial Consortium has specified policies and guidance, based on W3C and IETF standards, for naming authority best-practices here: **http://www.opengeospatial.org/standards/na**. Note specifically, the policy document titled "OGC-NA Name type specification - specification elements"(10-103). The generic scheme for identifying persistent names for online resources (e.g., service type identifiers) follows this pattern: http://{namingAuthority}/spec/{specName}/{version}

- **Esri REST Map Service**
  - Specification Identifier**: http://www.geoplatform.gov/spec/esri-map-rest –** a service compliant with the Esri ArcGIS Map Server REST API.
- **Esri REST Image Service**
  - Specification Identifier**: http://www.geoplatform.gov/spec/esri-image-rest –** a service compliant with the Esri ArcGIS Image Server REST API.
- **Esri REST Feature Service**
  - Specification Identifier**: http://www.geoplatform.gov/spec/esri-feature-rest –** a service compliant with the Esri ArcGIS Feature Server REST API.

For more information about identities for Esri REST API Specifications, see https://www.geoplatform.gov/spec.

***Additional identifiers can be added to the domain as relevant** Please verify additional identifiers with the Geospatial Platform Team (GeospatialPlatformTeam@fgdc.gov before implementing.*

3. The designated title for the service.

4. A description of the service including the service specification author, e.g. OGC, Esri, and relevant information about the content, function, intended audience, etc.

ISO metadata developers are encouraged to document their geospatial web service(s) using a stand-alone *Service Identification* record. By doing so, the metadata for datasets hosted by the service, current and future, can be linked to the same Service Identification Record and information about the service is maintained and updated in a central location. However, it may be more expedient in the near term to add the service information to existing dataset metadata records **within the *CI_OnlineResource*** component of any of the following:

- Service Identification (within the dataset record)
- Data Identification
- Distribution Transfer Option
- Distribution Distributor

Metadata Publication Best Practices

Provide Collection-level Metadata

Collections and series are comprised of data resources that share similar, homogenous, content but might vary in terms of content date or geographic extent. Examples include orthoimagery, elevation points, hyrdrography, and land cover. Collection-level metadata is useful in guiding users toward specific data resources.

When documenting a resource collection:

1. Create a metadata record for the entire collection (parent) that provides a(n):
   - *Title* that references the resource as a collection or series
   - *Identifier* that uniquely identifies the data collection
   - *Abstract* that identifies and describes the resource as a collection or series, e.g. 'This is a collection-level metadata record'.
   - *Geographic Extent* and *Temporal Extent* for the complete collection
   - *Online Linkage* (ISO) or *Network Address* (CSDGM) to websites that describes the collection and provides access to individual members. If using ISO metadata, indicate the appropriate *Online Function Code*, e.g. 'information', 'download', 'search', etc. If using CSDGM metadata, include the
2. Create metadata records for the individual collection members (children) and subset compilations, as feasible. The collection-level metadata can be transformed to an individual record by editing the:
   - *Title* that references the resource as a member of a collection or series and indicates the specific geography and/or temporal extent
   - *Identifier* that uniquely identifies the collection member
   - *Abstract* that identifies and describes the resource as a member of the collection of series, e.g. 'This resource is a member of a collection'.
   - *Geographic Extent* and/or *Temporal Extent* for the individual member.
   - *Online Linkage* that provides direct access to the individual member.

   In addition, the metadata record for an individual member of a resource collection should include reference to the larger collection
   - If using ISO 19115, include the:
     - *Title* of the resource collection MD_Identification.aggregationInfo>Md_AggregationInfo.aggregateDataSetName>Citation.title
     - *Type of Association* MD_Identification.aggregationInfo>Md_AggregationInfo.associationType>DS_AssociationTypeCode, e.g. largerWorkCitation, partOfSeamlessDatabase, collectiveTitle, series
     - *Identifier* for the resource collection MD_Identification.aggregationInfo>Md_AggregationInfo.aggregateDataSetIdentifier>Citation.identifier
   - If using ISO 19115-1, include the:
     - *Title* of the resource collection MD_Identification.associatedResource>Md_AssociatedResource.name>Citation.title
     - *Type of Association* MD_Identification.associatedResource>Md_AssociatedResource.associationType>DS_AssociationTypeCode, e.g. largerWorkCitation, partOfSeamlessDatabase, collectiveTitle, series
     - *Identifier* for the resource collection MD_Identification.associatedResource>Md_AssociatedResource.name>Citation.identifier
   - If using CSDGM, include:
     - *Title* of the resource collection Identification>Citation>Larger Work Citation>Citation.title
     - *Identifier* for the resource collection

Identification>Citation>Larger Work Citation>Citation.otherCitationDetails using the following format: 'identifier=(*identifier value*)

Once metadata is created for both the collection and collection members, it is important to organize and place the metadata in a manner that enables Data.gov to harvest collection-member metadata and the collection-level metadata without introducing duplicates.

At this time, Data.gov requires:

1. placing all the member metadata records in a single WAF *separate from the collection-level metadata record*
2. registering the WAF that includes the member metadata records as a "WAF Homogeneous Collection" Harvest Source in catalog.data.gov
3. when creating the "WAF Homogeneous Collection", providing agency-specific URLs for the following Harvest Source registration elements is key:
   - ***URL*** – provide a link to the WAF that contains the member metadata records
   - ***Collection Metadata URL*** – provide a link to the collection-level metadata record

By organizing and registering the collection-level and member metadata records in this manner, Data.gov and the Geospatial Platform are able to point users to member metadata records upon discovery of the collection-level metadata record and greatly facilitate user access to the resource that best meets their information needs.

In the near future, however, Data.gov expects to utilize the metadata record identifiers to connect data resources with collections to which they belong and separate harvest folder will no longer be require. This guidance document will be updated at that time.

Don't Publish Metadata for Resources Produced by Others

Metadata should be published to Data.gov by the agency that is responsible for the resource. If metadata for the same resource is duplicated, it can result in conflicting information about the resource, confusion about the resource authority, and the unintended use of derived resources. When committing resources to Data.gov, a publisher is asserting that the resources conform to the data quality guidelines of the publishing organization and that they are authors of such data.

If an organization modifies a resource, the *Title*, *Abstract*, and *Lineage* of the metadata record for the modified resource should make clear the modification and attribute both the source resource and the authority responsible for the resource.

In some cases, publishers may serve as a metadata clearinghouse for other organizations, such as a federal Enterprise Data Inventory (EDI) or state metadata catalog. These publishers are encouraged to publish to Data.gov if the metadata records within their catalog are unique and clearly cite the authority responsible for the resource. If the publisher cannot assure that the records are unique, they are also encouraged to publish if they assure that the *Title and Abstract* for each metadata records includes the publisher's name, e.g. 'State of Oregon Metadata Clearinghouse – National Hydrography Dataset (NHD)…'